Association for Information Systems
AIS Electronic Library (AISeL)

CONF-IRM 2016 Proceedings

International Conference on Information Resources Management (CONF-IRM)

2016

Mechanisms that are impacting the Kenya Open Data Initiative

Paul Mungai
University of Cape Town, paul.mungai@pearson.com

Jean-Paul Van Belle
University of Cape Town, Jean-Paul.VanBelle@uct.ac.za

Joseph Sevilla
Strathmore University, joe@strathmore.edu

44. Mechanisms that are impacting the Kenya Open Data Initiative

Paul Mungai
University of Cape Town
paul.mungai@pearson.com.

Jean-Paul Van Belle
University of Cape Town
Jean-Paul.VanBelle@uct.ac.za

Joseph Sevilla
Strathmore University
com za joe@strathmore.edu

Abstract
Open Government Data (OGD) has become a topic of prominence during the last decade. However, most governments have not realised the desired outcomes from OGD, which implies that the envisaged value streams have not been realised. In order to help address this shortcoming, this study aims at identifying the candidate causal mechanisms that are impacting on OGD initiatives. This will be achieved through the lens of critical realism. This will assist implementers of OGD to formulate policies and structures that will help ensure that the initiative is sustainable and capable of achieving the set objectives and goals. Given the inadequacy of current literature on causal mechanisms that impact on OGD initiatives, it will also contribute to the existing OGD literature, mainly through the case studies and the causal mechanisms that will emerge.

Keywords
Open Government Data, Critical Realism, Mechanisms, Kenya Open Data Initiative

1. Introduction
The research question for this study is what are the candidate causal mechanisms that impact on the Kenya Open Data Initiative (KODI)? These mechanisms will be identified through the lens of critical realism, which will be described in the section that follows. As a note, this study will consider both enabling and disabling mechanisms. The following paragraphs describe the open government data phenomenon, including its potential value, and challenges once implemented.

Open government data (OGD) consists of three major aspects; data, government data and open data. Data refers to any information or recordings that is stored electronically, which implies that data that is of public interest should be converted to electronic format to the best feasible extent. Government data refers to all data and information generated by a public institution. Open data refers to "data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike" (Open Knowledge Foundation, 2012). In december 2007, the open government data working group formulated a set of eight principles aimed at guiding the process of making government data open. These include; complete – all data that does not violate privacy or security should be made available, primary – should be collected at source with high levels of granurality, timely – should be published with minimal delays, accessible – should be available to

the widest range of users, machine processable, non-discrimatory – available to anyone, non-proprietary, and license-free - not subject to any copyright, patent, trademark or trade secret regulation though this comes with an exception that reasonable security, privacy and privilege restrictions may be allowed (Open Knowledge Foundation, 2012).

1.1 OGD Value

OGD has gained prominence over the years following it's perceived value, which includes it's ability to; foster transparency, improve accountability, satisfy legal obligations, improve public-government participation and collaboration, foster responsiveness and democratic control, increase public awareness of government programmes and activities, and foster innovation, efficiency and effectiveness in government services (Böhm et al., 2012; Hoxha & Brahaj, 2011; Robinson, Harlan, Zeller, & Felten, 2009; Shadbolt, Hara, Berners-lee, Gibbins, & Glaser, 2007; Ubaldi, 2013).

The stated value of OGD is in form of services, which implies that OGD has no value in itself except when it is offered as a service to the public (M. Janssen, Charalabidis, & Zuiderwijk, 2012). However, it is important to note that these value streams are not easy to realize, as witnessed in the countries that have implemented OGD (Shadbolt et al., 2007). Majority of the countries that have implemented OGD are in Europe, North America and Australia (K. Janssen, 2011; Shadbolt et al., 2007; Ubaldi, 2013; Yu & Robinson, 2012).

In collaboration with the government, citizens can help actualize some of the potential value by providing insights to government using the availed OGD. They can also use OGD to provide new services that are aimed at value addition. Overall, one of the main objectives/outcome is to improve decision making of both government and the citizens, which can be achieved when the right data is availed (Dawes, 2005; Shadbolt et al., 2007; Ubaldi, 2013).

1.2 OGD Challenges

OGD faces several challenges, which affect the possibility of realising the potential value of OGD. These include: disclosure policies which limits OGD provisioning; copyright which creates contention on who owns government data; poor data quality and management practices which increases the cost of converting the data in machine readable format; enormous and discrete nature of government data that requires extra effort and cost when transforming it to OGD; finding a dedicated government agency that solicits datasets from other government agencies; increasing public interest and awareness of OGD that includes public servants, citizens and the private sector (Ubaldi, 2013).

In some countries, the very government agents who are meant to assist in the process of curating and publishing OGD become the stumbling blocks through resistance to change. In Cameroon for instance, government officials at the Ministry of Public Service and Administrative Reform refused to use the e-government system whose

main aim was to improve transparency and accountability following a rise in corruption and the number of ghost workers. This failure resulted from a disconnect between technology and the social context, which needs to be bidirectional (Heeks, 2005). In other cases, government agents "consider public information their own property and not of the citizen" leading to resistance in releasing information/documents (Meijer, 2012). An example of this was observed in Vienna, the capital city of Austria. The city is decentralized and comprises of several heads of departments who report to the city directorate on administrative matters. The city directorate issued a regulation on OGD, which requires all departments to release open data. However, like many other regulations issued by the city directorate, this directive was not adhered to, and departments retained control on what data to release and in what granularity (Parycek, Höchtl, & Ginner, 2014). This implies that internal interests affect OGD provisioning.

Another challenge is lack of governance structures that specifically address OGD, which implies that e-Government structures are not ideal for OGD. This is based on the vivid and fast changing nature of OGD, which may not be supported by the existing structures, such as the approval processes. Austria could be emulated in this regard, where a sleek governance structure was formulated to address this. Following approval, it was cascaded downwards to the provinces, cities and municipalities (Parycek et al., 2014).

Another challenge to OGD provisioning emanates from the factors that complicate access and integration. This follows the complex nature of OGD given its size, schematic heterogeneity, quality variations and lack of consistency (Böhm et al., 2012; Hoxha & Brahaj, 2011). This results from the fact that government has multiple agencies, which follow different standards of data presentation and also, the fact that these agencies produce different types of data, which call for different presentation styles creating a challenge for uniformity. There is also the lack of meta-data, which would assist in describing the data (Hoxha & Brahaj, 2011).

Related to the complexity of OGD, another challenge arises when government attempts to structure and publish processed data (Robinson et al., 2009). This often arises when the complexity that comes with the heterogeneous nature of OGD is ignored. It is difficult to develop sites that address the needs of all citizens. This is partly because there is lack of insight on their perspectives and needs (Janssen et al., 2012). To address this, government should focus on developing infrastructure that is capable of presenting the underlying OGD in open, structured and machine-readable format. This should not imply that private entities will understand, interpret and present this data correctly the first time, but it is believed that they are more capable of exploring more approaches faster and find solutions faster than the government would. This is partly because private entities deal with specific data sets while government would have to do this for all the data sets. They should also use open standards such as RSS (Really Simple Syndication) to notify users whenever new data is made available (Robinson et al., 2009). This implies the need to ensure that OGD is fit for use. OGD cannot be universally fit for all users since requirements vary. This implies that users need to have access to good data descriptors that guide them in

deciding on the appropriateness of the data in question. This calls for the provisioning of metadata – data that describes data. Noting the importance of metadata, we can deduce that "good quality metadata is as important as the quality of the data itself"(Dawes & Helbig, 2010; Ubaldi, 2013).

2. Critical Realism

Compared to positivism, intepretivism and critical research, critical realism is a new entrant in social science research (Mingers, 2004; Orlikowski & Baroudi, 1991; Smith, 2006; Wynn & Williams, 2012). It brings in a new approach to research, following it's ability to identify the underlying mechanisms between indeterminate events and interactions, and provide in-depth causal explanations that assist in formulating systems-oriented information systems theories (Wynn & Williams, 2012).

Critical realism is based on the notion that events should be investigated at the level of generative mechanism that occur in the real domain, not at the level of constant conjunction for regular events since establishing a constant conjunctive relationship is not sufficient (Easton, 2010; Mingers, 2002; Smith, 2006). These mechanisms could be likened to the connections between variables, from which outcomes emerge (Fox, 2009). Events are selected for investigation based on their ability to have causal effect on the world. This differs from empiricism that selects events based on perceptability - the notion that only that which can be perceived can exist (Easton, 2010; Mingers, 2002).

Events can be investigated at either the empirical, actual or real domain, which are defined as follows: Empirical domain contains events that are observed or experienced; Actual domain contains events that do or do not occur, including those in the empirical domain; Real domain contains the whole of reality that includes mechanisms, events and experiences (Mingers, 2002). Events in the real and actual domain may not be observable at all or even when they are, observers may understand them quite differently (Easton, 2010; Mingers, 2002). This is likely to imply that events can only be observed in the empirical domain. However, this is not the case, but simply that events may not always be capable of being observed in the real or actual domain, thus creating a need for experimentation. Also, the conditions established by the observer during experimentation do not cause the results, which are dependent on causal laws at play (Easton, 2010; Mingers, 2002). Following this understanding of the various domains, critical realism suggests using the empirical domain during investigation (Easton, 2010; Mingers, 2002).

The empirical domain could be likened to the tip of an iceberg, where only a part is visible, and it is that which we observe. However, this should not imply that what is invisible is non-existent or unconnected to the visible (Easton, 2010). This analogy leads to a fundamental epistemological assumption in critical realism, that no observation is infallible (Easton, 2010; Mingers, 2004). This follows the realization that, under the empirical domain, it is unlikely to make observations that will result in full understanding of the social situation in question. Also, that there is no definitive

criteria to judge the "truth" of a particular explanation. Therefore, there is need for the observer to collect sufficient data that will aid in distinguishing alternative explanations of the same or a similar social situation (Easton, 2010; Smith, 2006).

Critical realism has three main benefits to information systems research. First, it helps in transcending a number of inconsistencies between stated philosophical assumptions and the actual practice of information systems research, under both positivism and interpretivism. Second, it offers a way to address the rigor-relevance gap in management research following its approach to causal analysis through multi-method/triangulation and multilevel approaches. This implies that critical realism is not limited to the case study approach and also that it can support several methods and approaches within a single study. Third, it assists in identifying connections between technology implementations and their outcomes (Wynn & Williams, 2012).

3. Methodology

Critical realism aims at formulating causal explanations that explain the way things act and how they are capable of doing so in a socio-technical context. To achieve this, case study method is preferred among many critical realism researchers (Easton, 2010; Mingers, 2004; Smith, 2006; Wynn & Williams, 2012). This follows it's ability to study a phenomenon within one or a small number of social entities or situations within a real-life context using multiple sources of data, which comprise of complex structures that are difficult to access, and which cannot be studied outside the context of occurrence (Dube & Pare, 2003; Easton, 2010; Wynn & Williams, 2012).

In order to tease out and disentangle these complexities, it starts by identifying the research questions, followed by the case selection criteria, which includes the boundary definition. The selected case comprises of a single or manageable number of entities to obtain data (Dube & Pare, 2003; Easton, 2010; Eisenhardt, 1989; Kvale, 1996). Following this, data is collected using triangulation (mixed methods) (Dube & Pare, 2003; Wynn & Williams, 2012). It is important to note that the data collection instruments are guided by the candidate theories, which are formulated through induction from the information gathered from literature review. Through deduction, theories are then used to formulate the potential Context Mechanism Outcome (CMO) configurations. Observations/data collection is conducted based on these through retroduction with the aim of identifying the CMO configurations that occur with regularity in the case at hand (Ranmuthugala et al., 2011).

Once the data is obtained, the case is written iteratively, giving a holistic description of the observed entities, which provides causal explanations about the phenomena in question (Dube & Pare, 2003; Easton, 2010; Eisenhardt, 1989; Flyvbjerg, 2006). Once the case is complete, the program specification is created (Ranmuthugala et al., 2011). This is achieved through empirical corroboration, which entails reviewing, validating and refining the proposed theories and potential CMO configurations using the empirical observations made in the previous stage (Easton, 2010; Popper, 2014; Ranmuthugala et al., 2011; Wynn & Williams, 2012).

The following table describes how the methodology described above has applied for this study and the description and status of each where it applies.

Item	Description
Research question	What are the mechanisms that enable proper institutionalization of the Kenya Open Data Initiative?
Candidate theories/mechanisms	Formulated through induction from the information gathered from literature review **Status:** Complete
Potential Context Mechanism Outcome (CMO) configurations	Formulated through deduction from the candidate theories **Status:** Complete
Case Study	**Unit of analysis:** Kenya Open Data Initiative institutionalization **Approach based on** (Yin, 1994) • Use a single-holistic unit of analysis • Use the explanation building technique in answering how and why questions - in relation to mechanisms **Case selection criteria based on** (Yin, 1994) • Alignment of the case with the centrality of the institutionalization process • Willingness to provide required access for the study • Availability of diverse organizational actors leading to multiple perceptions of practice • Availability during the entire duration of study **Status:** Ongoing
Data collection/sources	**Mixed method approach:** semi-structured interviews, observation, and document review. **Target institutions:** Kenya ICT Authority, Kenya Bureau of Statistics, World Bank Kenya, Strathmore University – iLab Africa, Open Institute (Civil Society Organisation (CSO)), Development Initiatives (CSO), Code 4 Kenya, and Data Science Ltd. **Interviews**: Nineteen interviews have been conducted so far while ten are pending. **Documents**: Fifteen documents in the form of blogs, meeting minutes, publications on KODI, and newspaper articles. There are also system logs that demonstrate usage of the KODI portal and tweeter archives on discussions related to KODI. **Status:** Ongoing
Data Analysis	**Approach:** • Content and narrative analysis • Abduction and retroduction to propose and test potential mechanisms **Status:** Ongoing
Generative mechanisms	**Activities:** • Identify generative mechanisms from the case study. • Describe each mechanism based on case study data and institutionalization theories. • Describe the contextual factors that shape each of these mechanisms. • Categorise them into core, direct and supportive mechanisms. **Status:** Yet to commence.
Program specification	This entails reviewing, validating and refining the proposed theories and potential CMO configurations using the empirical observations made earlier. **Status:** Yet to commence.

Table 1: KODI Study Methodology

The pending interviews are focused on data fellows that were assigned to various government agencies to help on technical aspects of OGD data curation and publication. The data fellows are a product of the data fellows program, an initiative of the Kenya ICT Authority. It is aimed at strengthening the capacity of the host institutions to generate and publish data sets of public interest. The other focus is on staff in government agencies that act as champions in charge of institutionalizing the OGD initiative within their institutions. These interviews will be conducted in mid February, following an agreement with the ICT Authority in relation to facilitation and availability of staff in the various agencies. The aim of these interviews will be to further understand what internal policies have been formulated within these agencies to help them implement the national (external) OGD policy/directive.

4. KODI Candidate Causal Mechanisms

This study has been able to identify three candidate mechanisms during the inductive phase. This involved a systematic search and review of literature. The aim of this process was to identify the events, structure, context and outcomes of Government Open Data Initiatives. It also aimed at identifying the components of structure and the variations of contextual influences and candidate mechanisms. The findings of this phase, mainly the preliminary CMOs and candidate mechanisms helped in formulating interview questions. These mechanisms are described in the table below, which helps in describing the observable events, the people and systems involved or required for each, the conditions or pressures they are subjected to, and the desired outcomes.

Candidate Mechanisms	**Events**	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**
Law & Policy Reinforcement	- Obtain support from the President. - Establish and implement a legal framework and policies on right of access to information, confidentiality, exceptions to openness, and intellectual property rights. - Control publication and use of data using copyright laws and disclosure policies. - Protect government agents privacy. - Devolve decision-making. Allows stakeholders a stronger say in choices of government programs and services. It also supports proactive disclosure.	- President - ICT Cabinet Secretary - Devolution & Planning Cabinet Secretary - ICT Authority - Kenya National Bureau of Statistics - Government agencies - State corporations - Civil society	- Support from the President - Proactive disclosure policies - Devolved decision-making - Shared vision on KODI - Organizational cohesion – commitment from the entire government. - Availability of resources to implement formulated law and policy	- Political leadership in support of KODI - Adequate budget allocation - Implementation of formulated law and policy by government agencies and state corporations - Sustainability of KODI - Reduced data hoarding and release of tampered data
Coordination and Capacity Building	- Monitor & report on open data projects - Facilitate skill and resource acquisition - Create awareness on the essence of open data - Educate on OGD laws and policies - Educate on proactive disclosure of OGD. This entails releasing data without waiting for specific data requests from the public	- ICT Cabinet Secretary - Devolution & Planning Cabinet Secretary - ICT Authority - Government agencies - State corporations - Civil society	- Availability of skills and resources to empower staff - Adequate allocation of funds - Efficient expenditure approval channels - Ability to influence change in government agencies and state corporations	- Timely publication of open data - Increased corporation between government agencies and state corporations - Reinforced value for users - Improved public service delivery - Accurate reporting of open data projects
Advocacy	- Organize conferences & boot camps - Partner and engage with civil society - Conduct informative sessions aimed at: increasing public interest & preparedness; appreciating the value of crowd sourcing; changing the attitude of public officials on openness; ensuring stakeholder buy-in - Monitor progress on OGD projects	- ICT Cabinet Secretary - Devolution & Planning Cabinet Secretary - ICT Authority - Government agencies - State corporations - Civil society	- Ability to influence change in government agencies and state corporations - Availability of expertise and resources for advocacy	- Increased stakeholder awareness & support - Sustaining and strengthening the image of KODI - More OGD users - Increased public-private partnership models -Faster OGD bottleneck resolution

Table 2: KODI Candidate Causal Mechanisms

There were more mechanisms that had been identified initially. Some of these included efficiency, transparency, innovation, crowdsourcing, data quality, government commitment, stakeholder engagement and participation. However, they were disqualified based on the following reasons. Some were similar or related and were merged as a result. Others did not have sufficient content to describe the context and possible outcomes. The other reason is based on the concept of generative mechanism, which suggests that mechanisms need to interact together in certain ways and have feedback relations which result in observable events (Mingers, 2014). The next diagram describes the interaction between the three mechanisms, and includes the activities that are exchanged between them.

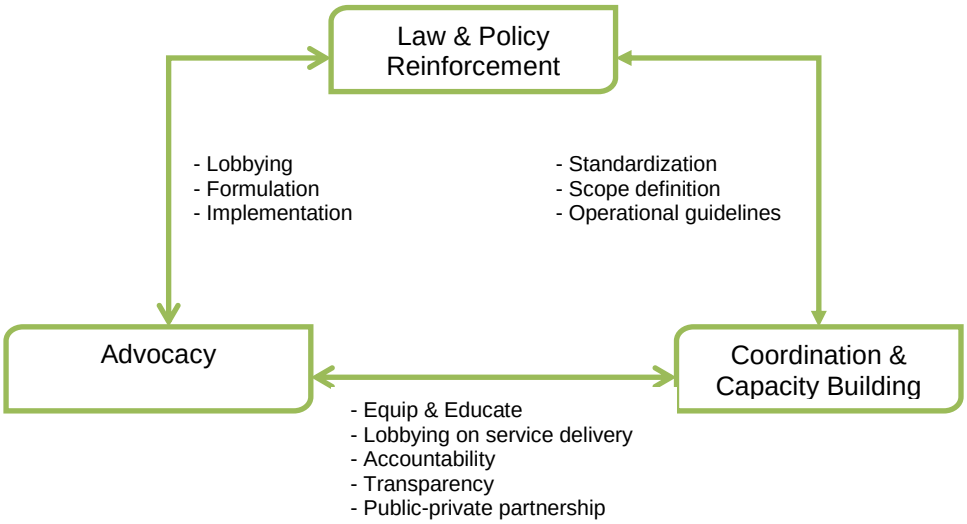

Figure 1: KODI Candidate Causal Mechanisms

5. Conclusions

This study has helped identify the candidate mechanisms that impact on open government data initiatives. This is useful in guiding ongoing or upcoming initiatives on what to invest in, facilitate or enable especially in regard to resources, policies and procedures. This study has also applied critical realism in an information systems initiative, and tried to demonstrate how such research should be conducted. Though this is not meant to be prescriptive, it creates an opportunity for discussion on what critical realism assumptions needs to be factored in such a study and how such should be implemented. This helps address one of the main challenges of critical realism in the field of information systems, which is to understanding how to apply critical realism in understanding the empirical domain in an information systems context.

References

Böhm, C., Freitag, M., Heise, A., Lehmann, C., Mascher, A., Naumann, F., … Schmidt, M. (2012). GovWILD : Integrating Open Government Data for Transparency. In *21st international conference companion on World Wide Web. ACM* (pp. 3–6).

Dawes, S. S. (2005). A Realistic Look at Open Data. *Center for Technology in Government, University at Albany/SUNY*.

Dawes, S. S., & Helbig, N. (2010). Information Strategies for Open Government : Challenges and Prospects for Deriving Public Value from Government Transparency. *Electronic Government*, 50–60.

Dube, L., & Pare, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.

Easton, G. (2010). Critical realism in case study research. *Industrial Marketing Management*, *39*(1), 118–128. doi:10.1016/j.indmarman.2008.06.004

Eisenhardt, M. K. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532–550. Retrieved from http://www.jstor.org/stable/258557

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, *12*(2), 219–245. doi:10.1177/1077800405284363

Fox, S. (2009). Applying critical realism to information and communication technologies: a case study. *Construction Management and Economics*, *27*(5), 465–472. doi:10.1080/01446190902842292

Heeks, R. (2005). e-Government as a Carrier of Context. *Journal of Public Policy*, *25*(1), 51–74. doi:10.1017/S0143814X05000206

Hoxha, J., & Brahaj, A. (2011). Open Government Data on the Web: A Semantic Approach. *2011 International Conference on Emerging Intelligent Data and Web Technologies*, 107–113. doi:10.1109/EIDWT.2011.24

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, *28*(4), 446–456. doi:10.1016/j.giq.2011.01.004

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. doi:10.1080/10580530.2012.716740

Kvale, S. (1996). The 1000-Page Question. *Qualitative Inquiry*, *2*(3), 275–284.

Meijer, A. (2012). Understanding the Complex Dynamics of Transparency Albert. *Public Administration Review*, *73*(3), 429–439.

Mingers, J. (2002). Real-izing Information Systems: Critical Realism as an Underpinning Philosophy for Information Systems. In *International Conference on Information Systems* (pp. 294–303).

Mingers, J. (2004). Real-izing information systems: critical realism as an underpinning philosophy for information systems. *Information and Organization*, *14*(2), 87–103. doi:10.1016/j.infoandorg.2003.06.001

Mingers, J. (2014). *Systems Thinking, Critical Realism and Philosophy: A Confluence of Ideas*. Routledge.

Open Knowledge Foundation. (2012). The Open Data Handbook. Retrieved January 5, 2015, from http://opendatahandbook.org/guide/en/

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations : Research Approaches and Assumptions. *Information Systems Research*, *2*(1), 1–28. Retrieved from http://www.jstor.org/stable/23010611

Parycek, P., Höchtl, J., & Ginner, M. (2014). Open Government Data Implementation Evaluation. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 13–14. doi:10.4067/S0718-18762014000200007

Popper, K. (2014). Corroboration, or how a theory stands up to tests. In *The logic of scientific discovery* (pp. 248–283). Routledge.

Ranmuthugala, G., Cunningham, F. C., Plumb, J. J., Long, J., Georgiou, A., Westbrook, J. I., & Braithwaite, J. (2011). A realist evaluation of the role of communities of practice in changing healthcare practice. *Implementation Science, 6*, 49. doi:10.1186/1748-5908-6-49

Robinson, D., Harlan, Y., Zeller, W. P., & Felten, E. W. (2009). Government data and the invisible hand. *Yale JL & Tech*, *11*(1).

Shadbolt, N., Hara, K. O., Berners-lee, T., Gibbins, N., & Glaser, H. (2007). Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*, *27*(3), 2–10.

Smith, M. L. (2006). Overcoming theory-practice inconsistencies: Critical realism and information systems research. *Information and Organization*, *16*(3), 191–211. doi:10.1016/j.infoandorg.2005.10.003

Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives* (No. 22). Retrieved from http://dx.doi.org/10.1787/5k46bj4f03s7-en

Wynn, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly*, *36*(3), 787–810.

Yin, R. (1994). *Case study research: Design and methods*. Beverly Hills.

Yu, H., & Robinson, D. G. (2012). *The New Ambiguity of "Open Government"* (Vol. 178). Retrieved from http://ssrn.com/abstract=2012489